BioNanoScience (2018) 8:384–389
https://doi.org/10.1007/s12668-017-0499-7

Prediction of Aromatic Hydroxylation Sites for Human CYP1A2 Substrates Using Condensed Graph of Reactions

T. I. Madzhidov[1] · A. A. Khakimova[1] · R. I. Nugmanov[1] · C. Muller[2] · G. Marcou[2] · A. Varnek[1,2]

Published online: 16 January 2018

Abstract
In this paper, support vector machine and condensed graph of reaction (CGR) approaches have been used to predict the regioselectivity of aromatic hydroxylation for human CYP1A2 substrates. Experimental data on aromatic hydroxylation for human cytochrome CYP1A2 (observed molecular or "real" transformations) used in the modeling were extracted from the Metabolite database and the XenoSite database. In addition, all potential but unobserved ("unreal") transformations were generated. The dataset containing "real" and "unreal" transformations was converted into an ensemble of CGRs representing pseudomolecules with conventional (single, double, aromatic, etc.) bonds and dynamic bonds characterizing chemical transformations. ISIDA fragment descriptors generated for CGRs were used for the modeling. The models have been validated in three times repeated fivefold cross-validation on the training set and then on an external set. The final model was constructed by consensus over models built on different descriptors sets. Predictive performance of our model on the external test set was similar to that of XenoSite and Way2Drug tools. Unlike previously used atom labeling-based approaches, the proposed CGR-based representation of metabolic transformations could be applied to different types of reactions catalyzed by the same enzyme and therefore, it is more suitable for automatized handling of metabolic data.

Keywords CYP1A2 · Aromatic hydroxylation · Support vector machine (SVM) · Condensed graph of reaction (CGR)

1 Introduction

Computational assessment of adsorption, distribution, metabolism, excretion, and toxicity (ADMET) properties [1] can significantly reduce costs of drug discovery process. Metabolic fate of xenobiotics in the human organism is one of the most difficult properties to predict. The point is that substrate oxidation could be catalyzed successively by different enzymes (belonging mostly to P450 cytochrome family). In phase I metabolism stage, molecules are oxidized and converted into their hydroxylated derivatives. These oxidations are often followed by conjugation reactions at phase II metabolism stage. Due to these transformations, the compounds become more polar (i.e., more soluble in water) which facilitates their excretion. Metabolic transformations may also modulate biological activities or toxicity of the ingested compounds. Thus, in order to anticipate undesired side effects, it is important to predict metabolites which could be formed.

Generally, computational studies [2, 3] reported so far focus on two main problems: (*i*) assessments of the specificity of a molecule toward different isoforms of cytochrome P450 and (*ii*) prediction of the most likely metabolic labile site(s). Some computational studies (docking, molecular interaction fields) [4–8] require explicit information about enzyme structure, whereas the others (QM calculations, QSAR, pharmacophore) [9–15] use only substrate structure. Up to now, reported ligand-based predictions outperform related structure-based ones. Thus, Sheridan et al. [16] developed QSAR models combining structural and physical property descriptors for predicting CYP3A4, 2D6, and 2D9 regioselectivity. They demonstrated that related QSAR models performed similarly to mechanism-based MetaSite approach. Swamidass et al. [17–19] reported a series of models predicting different metabolic transformations: the products of hydroxylation by cytochrome P450 [17], metabolic epoxidation centers [18], quinone formation [19], and metabolic transformation by UDP-glucuronosyltransferase

✉ T. I. Madzhidov
timur.madzhidov@kpfu.ru

✉ A. Varnek
varnek@unistra.fr

[1] Kazan Federal University, 18 Kremlyovskaya Str., Kazan, Russia 420008

[2] Université de Strasbourg, 4 rue B. Pascal, 67000 Strasbourg, France